Efficient Monte Carlo Methods
for the Potts Model at Low Temperature

Mehdi Molkaraie
Universitat Pompeu Fabra
08018 Barcelona, Spain
mehdi.molkaraie@upf.edu

Vicenç Gómez
Universitat Pompeu Fabra
08018 Barcelona, Spain
vicen.gomez@upf.edu

Abstract—We consider the problem of estimating the partition function of the ferromagnetic q-state Potts model. We propose an importance sampling algorithm in the dual of the normal factor graph representing the model. The algorithm can efficiently compute an estimate of the partition function when the coupling parameters of the model are strong (corresponding to models at low temperature) or when the model contains a mixture of strong and weak couplings. We show that, in this setting, the proposed algorithm significantly outperforms the state of the art methods.

I. INTRODUCTION

The Potts model [1] plays an important role in many areas, including statistical physics [2], image processing [3], and graph theory [4]. The fundamental quantity of interest in the model is the *partition function*. For example, in Bayesian model selection, the partition function represents the marginal likelihood (the evidence) [5, Chapter 5], in statistical physics, the internal energy follows from differentiating the log partition function [6, Chapter 2], and in deep belief networks, estimates of this quantity are needed in maximum likelihood learning [7]. In general, the exact computation of the partition function is intractable, as it requires a sum with an exponential number of terms. Although the planar Ising model (Potts with binary variables) without an external field is not difficult [8], the general 2D q-state Potts model without an external field is already computationally hard[1] [9].

In complex models, quantities of interest can be estimated via Monte Carlo (MC) methods (see, e.g., [10, 11, 12]). In the low-temperature regime, however, MC methods based on single spin-flips suffer from critical slowing down and have erratic convergence [13, 11]. More advanced MC methods (e.g., nested sampling [14] and the Swendsen-Wang algorithm [15]) require sampling from a large sequence of constraints or intermediate distributions at different temperatures to estimate the partition function.

We represent the Potts model using graphical models defined in terms of Forney/normal factor graphs (NFG) [16]. The partition function of an NFG is related to the partition function of the dual NFG via the normal factor graph duality theorem [17, 18]. In the models that we study here, this

theorem states that the two partition functions are equal up to scale. In [19], the authors showed that for the nearest-neighbor 2D Ising model, at low temperature, MC methods converge faster in the dual NFG than in the primal NFG. MC methods based on uniform and Gibbs sampling in the dual domain were proposed in [19, 20] to estimate the partition function.

In this paper, we extend the previous results to propose an importance sampling algorithm to estimate the partition function of the nearest-neighbor 2D ferromagnetic q-state Potts model. The proposed algorithm also operates in the dual NFG of the model. Our analytical results prove that we can obtain very accurate estimates of the partition function, when the coupling parameters on only a subset of the edges (which forms a spanning tree) are strong. Our numerical results show that, in various settings, our algorithm improves upon the state of the art MC methods in the dual domain and deterministic methos in the primal domain by more than an order of magnitude.

The paper is organized as follows. In Section II, we review the Potts model and its graphical model representation in terms of NFGs. Dual NFGs and the normal factor graph duality theorem are discussed in Section III. The dual NFG of the Potts model is used in Section IV to describe the proposed MC methods. Numerical experiments are reported in Section V.

II. THE 2D POTTS MODEL

Let $X_1, X_2, \ldots, X_N$ be a collection of discrete random variables arranged on the sites of a 2D grid, as illustrated in Fig. 1, where interactions are restricted to adjacent (nearest-neighbor) variables. Suppose that each random variable takes on values in a finite alphabet $\mathcal{X}$, which in this context is equal to the abelian group $\mathbb{Z}/q\mathbb{Z} = \{0, 1, \ldots, q-1\}$. Here, q is an integer satisfying $q > 2$.

Let x_i represent a possible realization of X_i, $\mathbf{x}$ stand for a configuration $(x_1, x_2, \ldots, x_N)$, and $\mathbf{X}$ stand for $(X_1, X_2, \ldots, X_N)$. A real coupling parameter $J_{k,\ell}$ is associated with adjacent variables (x_k, x_ℓ).

We start with the Potts model without an external field (the Potts model in an external field is discussed in Appendix I). The energy of a configuration $\mathbf{x}$ is given by the Hamiltonian [2]

$$\mathcal{H}(\mathbf{x}) = -\sum_{(k, \ell) \in \mathcal{B}} J_{k,\ell} \cdot [x_k = x_\ell] \qquad (1)$$

---

[1]There are, however, a few exactly solved cases, including the one-dimensional (1D) Potts model [6, Chapter 5], and in thermodynamic limits, for antiferromagnetic Potts models at zero temperature, the two-dimensional (2D) 3-state and the triangular 4-state models [2, Chapter 12].

Here, $\mathcal{B}$ contains all the unordered pairs (bonds) (k, ℓ) with non-zero interactions, and $[\cdot]$ denotes the Iverson bracket [21], which evaluates to 1 if the condition in the bracket is satisfied and to 0 otherwise. The parameter $J_{k,\ell}$ controls the strength of the interaction between (x_k, x_ℓ). In this paper, we concentrate on ferromagnetic Potts models, characterized by $J_{k,\ell} \geq 0$ for each $(k, \ell) \in \mathcal{B}$.

The probability that the model is in configuration $\mathbf{x}$ is given by the Boltzmann distribution [2]

$$p_{\mathrm{B}}(\mathbf{x}) = \frac{e^{-\mathcal{H}(\mathbf{x})/T}}{Z} \tag{2}$$

Here, the normalization constant Z is the partition function $Z = \sum_{\mathbf{x} \in \mathcal{X}^N} e^{-\mathcal{H}(\mathbf{x})/T}$ and T denotes the temperature. In the rest of this paper, we assume $T = 1$. Hence, large values (resp. small values) of J correspond to models at low (resp. high) temperature.

For each adjacent pair (x_k, x_ℓ), let $\kappa_{k,\ell} : \mathcal{X}^2 \to \mathbb{R}_{\geq 0}$

$$\kappa_{k,\ell}(x_k, x_\ell) = e^{J_{k,\ell} \cdot [x_k = x_\ell]} \tag{3}$$

We can then define $f : \mathcal{X}^N \to \mathbb{R}_{\geq 0}$ as

$$f(\mathbf{x}) \triangleq \prod_{(k, \ell) \in \mathcal{B}} \kappa_{k,\ell}(x_k, x_\ell) \tag{4}$$

From (4), Z in (2) can be expressed as

$$Z = \sum_{\mathbf{x} \in \mathcal{X}^N} f(\mathbf{x}) \tag{5}$$

The factorization in (4) can be represented by an NFG, in which nodes and edges represent factors and variables, respectively. The edge representing variable x is connected to the node representing factor $\kappa(\cdot)$ if and only if x is an argument of $\kappa(\cdot)$. If a variable (an edge) appears in more than two factors, such a variable is replicated via an equality indicator factor [16, 22]. The equality indicator factors are denoted by $\Phi_=(\cdot)$, which impose the constraint that all their incident variables be equal.

The NFG for the factorization in (4) is shown in Fig. 1, where the boxes labeled "=" are equality indicator factors. For variables X_2, X_2', X_2'' in Fig. 1, the equality indicator factor is as

$$\Phi_=(x_2, x_2', x_2'') = [\, x_2 = x_2' = x_2'' \,] \tag{6}$$

which evaluates to 1 if all its arguments are equal, and to 0 otherwise.

Throughout this paper, boundary conditions are assumed to be periodic. Thus, each equality indicator factor has degree four and $|\mathcal{B}| = 2N$. For simplicity, periodic boundary conditions are not shown.

At high temperature (i.e., small J), the Boltzmann distribution (2) approaches the uniform distribution. To estimate Z in this case, MC methods in the primal NFG, as in Fig. 1, perform well. In more challenging situations (e.g., at low temperature where (2) is highly non-uniform), we propose an importance sampling algorithm in the dual NFG of the 2D Potts model to estimate Z.
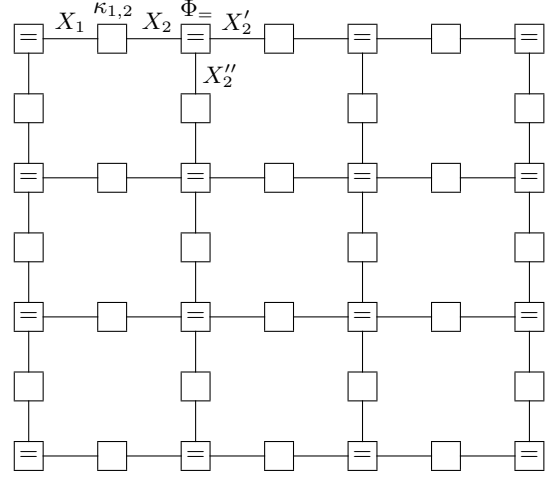

Fig. 1: The NFG of the 2D Potts model, where unlabeled boxes represent (3) and boxes containing "=" symbols are as in (6).
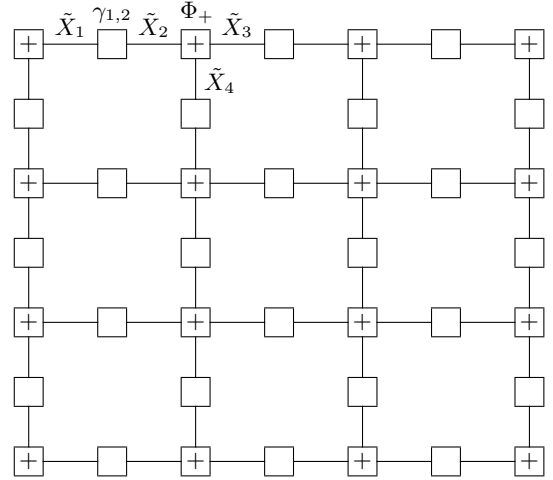

Fig. 2: The dual NFG of the 2D Potts model, where the unlabeled boxes represent factors (8) and boxes containing "+" symbols are as in (9).

III. THE DUAL NFG

We briefly discuss the normal factor graph duality theorem [16, 17, 18, 23], which relates Z to the partition function of the dual NFG, denoted by Z_{d}. We will also explain a procedure to obtain the dual NFG of the 2D Potts model. This procedure preserves the value of Z up to scale.

We will use the tilde symbol to denote variables in the dual domain. Note that $\tilde{X}$ also takes on values in $\mathcal{X}$. Here, $\gamma(\tilde{x}_1, \tilde{x}_2)$, the 2D discrete Fourier transform (DFT) of $\kappa(x_1, x_2)$, is as

$$\gamma(\tilde{x}_1, \tilde{x}_2) = \frac{1}{|\mathcal{X}|} \sum_{x_1 \in \mathcal{X}} \sum_{x_2 \in \mathcal{X}} \kappa(x_1, x_2) e^{-i 2\pi (x_1 \tilde{x}_1 + x_2 \tilde{x}_2)/|\mathcal{X}|} \tag{7}$$

where i is the unit imaginary number [24].

To obtain the dual of Fig. 1, we first replace each factor $\kappa_{k,\ell}(x_k, x_\ell)$ with its 2D DFT $\gamma_{k,\ell}(\tilde{x}_k, \tilde{x}_\ell)$, surrounded by two

factors $\mathcal{F}_{x_k}^{-1}$ and $\mathcal{F}_{x_\ell}^{-1}$. These complex-valued factors represent the inverse Fourier transform. We then merge each of these factors with the corresponding equality indicator factor, i.e., $\mathcal{F}_{x_k}^{-1}$ is merged with the equality indicator factor incident to the edge representing x_k, and $\mathcal{F}_{x_\ell}^{-1}$ with the equality indicator factor incident to x_ℓ.

After applying the above procedure to all the factors in Fig. 1, we obtain the dual NFG of the 2D Potts model, in which factors (3) are replaced by their 2D DFT, and equality indicator factors by their inverse DFT.

The 2D DFT of (3) has the following form

$$\gamma_{k,\ell}(\tilde{x}_k, \tilde{x}_\ell) = \begin{cases} e^{J_{k,\ell}} + q - 1, & \text{if } \tilde{x}_k = \tilde{x}_\ell = 0 \\ e^{J_{k,\ell}} - 1, & \text{if } \tilde{x}_k + \tilde{x}_\ell = 0 \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

The inverse DFT of an equality indicator factor (with degree four) is a mod_q indicator factor (also with degree four) – up to scale factor q. The mod_q indicator factors are denoted by $\Phi_+(\cdot)$. They impose the constraint that all their incident variables sum to zero (modulo q).

The dual NFG of the 2D Potts model is illustrated in Fig. 2, where the unlabeled boxes represent factors (8) and boxes labeled "+" are mod_q indicator factors. For variables $\tilde{X}_2, \tilde{X}_3, \tilde{X}_4$ in Fig. 2, the mod_q indicator factor is as

$$\Phi_+(\tilde{x}_2, \tilde{x}_3, \tilde{x}_4) = [\tilde{x}_2 + \tilde{x}_3 + \tilde{x}_4 = 0] \quad (9)$$

which evaluates to 1 if the sum of its arguments is zero, and to 0 otherwise.

After multiplying the N local scale factors that appear in the dualization procedure, we obtain

$$Z_{\mathrm{d}} = q^N Z \quad (10)$$

for more details, see [18, Theorem 2], [17].

In the following, we introduce a convenient modification to Fig. 2; the motivation will become apparent in Section IV. Note that $\gamma_{k,\ell}(\tilde{x}_k, \tilde{x}_\ell)$ is non-zero if and only if $\tilde{x}_k + \tilde{x}_\ell = 0$. In other words, (8) can equivalently be written as $\gamma_{k,\ell}(\tilde{x}_k - \tilde{x}_\ell)$. Therefore, by inserting a sign inverter in one of the edges incident to (8), it is possible to represent this factor using only one variable. For notational simplicity, we express this factor as $\gamma_k(\tilde{x}_k)$.

We can thus modify Fig. 2, to construct Fig. 3 with factors attached to each equality indicator factor

$$\gamma_k(\tilde{x}_k) = \begin{cases} e^{J_k} + q - 1, & \text{if } \tilde{x}_k = 0 \\ e^{J_k} - 1, & \text{otherwise,} \end{cases} \quad (11)$$

where J_k is the coupling parameter associated with each edge (bond) or "the bond strength".

The modified dual NFG is shown in Fig. 3, where the unlabeled boxes represent (11). In analogy with the logic NAND gate, the sign inverters are depicted by small circles attached to one side of the equality indicator factors; however, the choice can be made arbitrarily. See [16] for a similar approach in the context of non-binary codes (see also [25, 20] for the dual NFG of the Potts model).
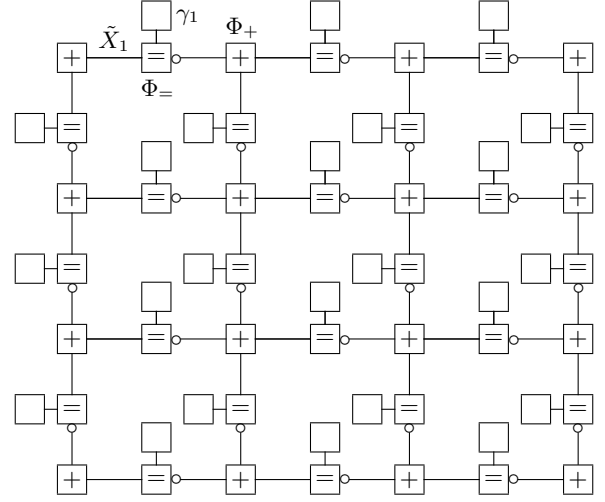

Fig. 3: Modified dual NFG of the 2D Potts model, where the unlabeled boxes represent factors (11) and the small circles denoted by "∘" are sign inverters.
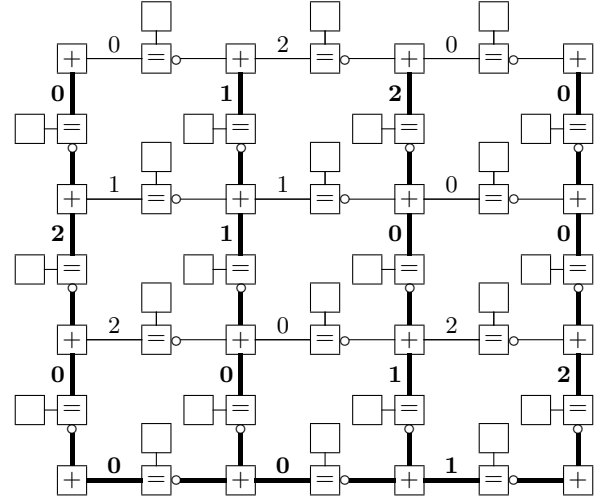

Fig. 4: A partitioning of the variables in Fig. 3, where the thin edges represent $\tilde{\mathbf{X}}_A$ and the remaining thick edges represent $\tilde{\mathbf{X}}_B$. Here, $\tilde{\mathbf{X}}_B$ is a linear combination of $\tilde{\mathbf{X}}_A$. A valid configuration in a 3-state Potts model is also illustrated.

In Section IV, we show that Fig. 3 can be used as an efficient calculational tool to compute MC estimates of Z_{d}, which can then be used to estimate Z via the normal factor graph duality theorem, cf. (10). In this paper, the focus is on ferromagnetic models, and as a result, factors (8) and (11) are nonnegative. We require all factors to be nonnegative as we need to define probability mass functions in Fig. 3 for our MC methods.

In the sequel, we will refer to Fig. 3 as the dual NFG or simply as the dual of the 2D Potts model.

IV. MONTE CARLO METHODS

We describe our MC methods in Fig. 3, where we partition the set of variables $\tilde{\mathbf{X}}$ into $\tilde{\mathbf{X}}_A$ and $\tilde{\mathbf{X}}_B$, with the condition that

$\tilde{\mathbf{X}}_B$ is a linear combination (involving mod_q indicator factors) of $\tilde{\mathbf{X}}_A$. In this set-up, $\tilde{\mathbf{X}}_B$ forms a spanning tree, and a valid configuration in Fig. 3 can be created by assigning values to $\tilde{\mathbf{X}}_A$, followed by computing/updating $\tilde{\mathbf{X}}_B$.

An example of such a partitioning, with a numerical example of a valid configuration in a 3-state Potts model, is illustrated in Fig. 4, where $\tilde{\mathbf{X}}_A$ is the set of all the variables associated with the thin edges (thin bonds) and $\tilde{\mathbf{X}}_B$ is the set of all the variables associated with the remaining thick edges. Accordingly, let $\mathcal{B}_A \subset \mathcal{B}$ contain the indices of the variables in $\tilde{\mathbf{X}}_A$, and $\mathcal{B}_B = \mathcal{B} - \mathcal{B}_A$.

The importance sampling algorithm works as follows. To draw $\tilde{\mathbf{x}}^{(\ell)}$ at iteration ℓ, we first draw a sample $\tilde{\mathbf{x}}_A^{(\ell)}$ according to an auxiliary (proposal) distribution in the dual NFG. We then update $\tilde{\mathbf{x}}_B^{(\ell)}$ to create a valid configuration $\tilde{\mathbf{x}}^{(\ell)} = (\tilde{\mathbf{x}}_A^{(\ell)}, \tilde{\mathbf{x}}_B^{(\ell)})$. Updating $\tilde{\mathbf{x}}_B^{(\ell)}$ is easy as $\tilde{\mathbf{x}}_B$ is a linear combination of $\tilde{\mathbf{x}}_A$.

Let us define

$$\Gamma_A(\tilde{\mathbf{x}}_A) \triangleq \prod_{k \in \mathcal{B}_A} \gamma_k(\tilde{x}_k) \tag{12}$$

$$\Gamma_B(\tilde{\mathbf{x}}_B) \triangleq \prod_{k \in \mathcal{B}_B} \gamma_k(\tilde{x}_k) \tag{13}$$

In our importance sampling algorithm, we use the following probability mass function as the auxiliary distribution

$$q_{\mathrm{d}}(\tilde{\mathbf{x}}_A) = \frac{\Gamma_A(\tilde{\mathbf{x}}_A)}{Z_{q_{\mathrm{d}}}}, \quad \forall \tilde{\mathbf{x}}_A \in \mathcal{X}^{|\mathcal{B}_A|} \tag{14}$$

where $|\mathcal{B}_A|$ denotes the cardinality of $\mathcal{B}_A$.

The partition function $Z_{q_{\mathrm{d}}}$ is analytically available as

$$Z_{q_{\mathrm{d}}} = \sum_{\tilde{\mathbf{x}}_A} \Gamma_A(\tilde{\mathbf{x}}_A)$$

$$= \prod_{k \in \mathcal{B}_A} \sum_{t=0}^{q-1} \gamma_k(t)$$

$$= q^{|\mathcal{B}_A|} \exp\Big(\sum_{k \in \mathcal{B}_A} J_k \Big) \tag{15}$$

It is also straightforward to draw *independent* samples $\tilde{\mathbf{x}}_A^{(1)}, \tilde{\mathbf{x}}_A^{(2)}, \ldots, \tilde{\mathbf{x}}_A^{(\ell)}, \ldots, \tilde{\mathbf{x}}_A^{(L)}$ according to $q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)$. The product form of (12) indicates that to draw $\tilde{\mathbf{x}}_A^{(\ell)}$ we can apply Algorithm 1, where in line 3, $\big(1 + (q-1)e^{-J_k}\big)/q$ is equal to $\gamma_k(0)/\sum_{t=1}^{q-1} \gamma_k(t)$.

Algorithm 1 Drawing $\tilde{\mathbf{x}}_A^{(\ell)}$ according to $q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)$

1: *draw* $u_1^{(\ell)}, u_2^{(\ell)}, \ldots, u_{|\mathcal{B}_A|}^{(\ell)} \overset{\mathrm{i.i.d.}}{\sim} \mathcal{U}[0,1]$
2: **for** $k = 1$ **to** $|\mathcal{B}_A|$ **do**
3:    **if** $u_k^{(\ell)} < \dfrac{1 + (q-1)e^{-J_k}}{q}$ **then**
4:       $\tilde{x}_{A,k}^{(\ell)} = 0$
5:    **else**
6:       *draw* $\tilde{x}_{A,k}^{(\ell)}$ *randomly in* $\{1, 2, \ldots, q-1\}$
7:    **end if**
8: **end for**

After drawing $\tilde{\mathbf{x}}_A^{(\ell)}$, we update $\tilde{\mathbf{x}}_B^{(\ell)}$ to create a valid configuration $\tilde{\mathbf{x}}^{(\ell)} = (\tilde{\mathbf{x}}_A^{(\ell)}, \tilde{\mathbf{x}}_B^{(\ell)})$. Algorithm 2 then uses these samples to estimate Z_{d}. It is easy to verify that $\hat{Z}_{\mathrm{IS}}$ in (17) is an unbiased estimator of Z_{d}. Indeed

$$\mathrm{E}_{q_{\mathrm{d}}}[\hat{Z}_{\mathrm{IS}}] = Z_{\mathrm{d}} \tag{16}$$

Algorithm 2 Estimating Z_{d}

1: **for** $\ell = 1$ **to** L **do**
2:    *draw* $\tilde{\mathbf{x}}_A^{(\ell)}$ *according to* $q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)$
3:    *update* $\tilde{\mathbf{x}}_B^{(\ell)}$
4: **end for**
5: *compute*

$$\hat{Z}_{\mathrm{IS}} = \frac{Z_{q_{\mathrm{d}}}}{L} \sum_{\ell=1}^{L} \Gamma_B(\tilde{\mathbf{x}}_B^{(\ell)}) \tag{17}$$

The accuracy of (17) depends on the fluctuations of $\Gamma_B(\tilde{\mathbf{x}}_B)$. If $\Gamma_B(\tilde{\mathbf{x}}_B)$ varies smoothly, $\hat{Z}_{\mathrm{IS}}$ will have a small variance. From (11) and (13), we expect to observe a small variance if J_k is large for $k \in \mathcal{B}_B$. See our analysis of the variance in Section IV-A.

An algorithm based on uniform sampling, as applied in [19, 20], can be obtained by drawing $\tilde{\mathbf{x}}_A^{(\ell)}$ uniformly and independently from $\mathcal{X}^{|\mathcal{B}_A|}$, i.e., according to

$$u_{\mathrm{d}}(\tilde{\mathbf{x}}_A) = \frac{1}{Z_{u_{\mathrm{d}}}}, \quad \forall \tilde{\mathbf{x}}_A \in \mathcal{X}^{|\mathcal{B}_A|} \tag{18}$$

where $Z_{u_{\mathrm{d}}} = |\mathcal{X}|^{|\mathcal{B}_A|}$.

After updating $\tilde{\mathbf{x}}_B^{(\ell)}$, we will use the created configurations $(\tilde{\mathbf{x}}_A^{(\ell)}, \tilde{\mathbf{x}}_B^{(\ell)})$ in

$$\hat{Z}_{\mathrm{Unif}} = \frac{Z_{u_{\mathrm{d}}}}{L} \sum_{\ell=1}^{L} \Gamma_A(\tilde{\mathbf{x}}_A^{(\ell)}) \Gamma_B(\tilde{\mathbf{x}}_B^{(\ell)}) \tag{19}$$

It is easy to verify that, $\mathrm{E}_{u_{\mathrm{d}}}[\hat{Z}_{\mathrm{Unif}}] = Z_{\mathrm{d}}$.

The performance of the uniform and importance sampling algorithms will be close in model at very low temperature (i.e., for large J_k). However, for a wider range of parameters, importance sampling outperforms uniform sampling (see Section V).

A. Analysis of the variance

We present a heuristic argument for the accuracy of our MC methods. Let us replace (11) by

$$\gamma_k(\tilde{x}_k) = (e^{J_k} + q - 1)\Big(\frac{e^{J_k} - 1}{e^{J_k} + q - 1} \Big)^{1 - [\tilde{x}_k = 0]} \tag{20}$$

$$\propto \Big(\frac{e^{J_k} - 1}{e^{J_k} + q - 1} \Big)^{1 - [\tilde{x}_k = 0]} \tag{21}$$

The required factor S to recover Z_{d} can be computed as $S = \prod_{k \in \mathcal{B}} (e^{J_k} + q - 1)$.

Note that (21) tends to a constant factor as J_k grows, which gives reasons for the fast convergence of the uniform sampling
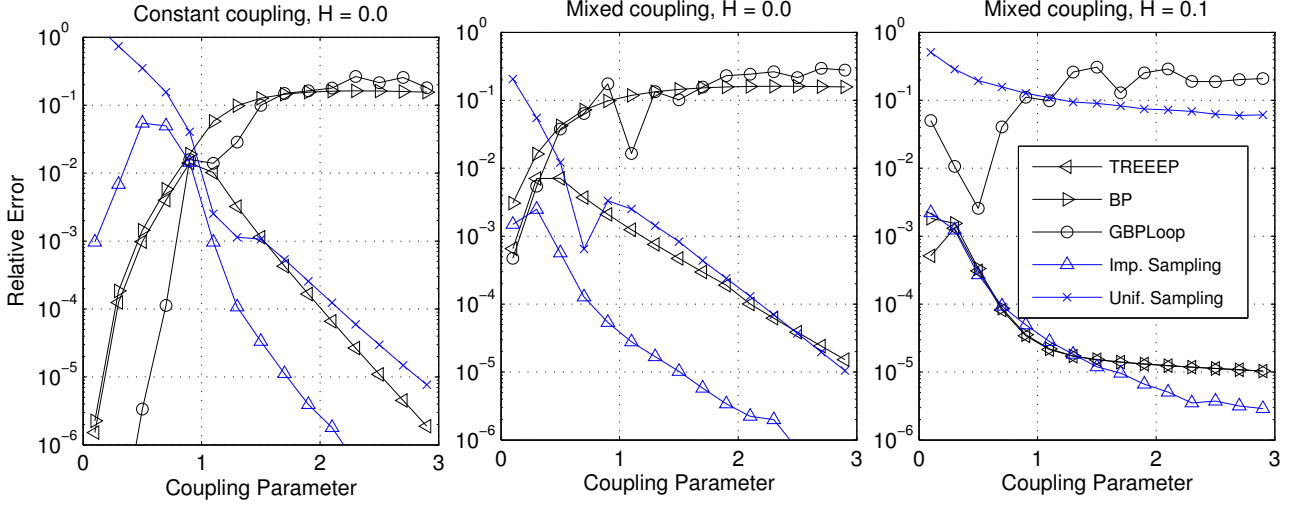
Fig. 5: The relative error $|\ln Z - \ln \hat{Z}|/\ln Z$ vs. the coupling parameter for an 8×8 3-state Potts model; **(left)** constant coupling with zero local field; **(middle)** mixed coupling, where we set $J_k = 2$ for $k \in \mathcal{B}_B$, and for $k \in \mathcal{B}_A$, we set J_k according to the value on the x-axis; **(right)** mixed couplings with a weak external field.

algorithm in this case (cf. (12), (13), and (19)), whereas the importance sampling scheme is guaranteed to have fast convergence when the coupling parameters associated with the thick edges are strong (cf. (13) and (17)). Indeed, in contrast to the convergence of MC methods in the primal NFG, convergence of uniform sampling improves as J_k becomes larger for each $k \in \mathcal{B}$. Convergence of importance sampling improves as J_k becomes larger for each $k \in \mathcal{B}_B$.

More rigorously, let us denote the global probability mass function in the dual factor graph by $p_{\mathrm{d}}(\cdot)$. Notice that $p_{\mathrm{d}}(\cdot)$ and $q_{\mathrm{d}}(\cdot)$ in (14) are both defined in the same configuration space $\mathcal{X}^{|\mathcal{B}_A|}$.

We can therefore write $p_{\mathrm{d}}(\cdot)$ as a function of $\tilde{\mathbf{x}}_A$

$$p_{\mathrm{d}}(\tilde{\mathbf{x}}_A) = \frac{\Gamma_A(\tilde{\mathbf{x}}_A)\Gamma_B(\tilde{\mathbf{x}}_B)}{Z_{\mathrm{d}}} \tag{22}$$

$$= \frac{Z_{q_{\mathrm{d}}}}{Z_{\mathrm{d}}} q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)\Gamma_B(\tilde{\mathbf{x}}_B), \quad \forall \tilde{\mathbf{x}}_A \in \mathcal{X}^{|\mathcal{B}_A|} \tag{23}$$

Following [26, 13], from (23) we express the variance of $\hat{Z}_{\mathrm{IS}}$ in (17) as

$$\frac{L}{Z_{\mathrm{d}}^2}\mathrm{Var}[\hat{Z}_{\mathrm{IS}}] = \frac{L}{Z_{\mathrm{d}}^2}\left(\mathrm{E}\left[\hat{Z}_{\mathrm{IS}}^2\right] - \left(\mathrm{E}\left[\hat{Z}_{\mathrm{IS}}\right]\right)^2\right) \tag{24}$$

$$= \left(\frac{Z_{q_{\mathrm{d}}}}{Z_{\mathrm{d}}}\right)^2 \mathrm{E}_{q_{\mathrm{d}}}[\Gamma_B^2(\tilde{\mathbf{X}}_B)] - 1 \tag{25}$$

$$= \sum_{\tilde{\mathbf{x}}_A} \frac{p_{\mathrm{d}}^2(\tilde{\mathbf{x}}_A)}{q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)} - 1 \tag{26}$$

$$= \chi^2\big(p_{\mathrm{d}}(\tilde{\mathbf{x}}_A), q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)\big) \tag{27}$$

where $\chi^2(\cdot, \cdot)$ denotes the chi-squared divergence, which is always non-negative, with equality to zero if and only if its two arguments are equal [27, Chapter 4].

For simplicity, let us assume that for $k \in \mathcal{B}_B$ the coupling parameters are constant, denoted by J_B. From (21),

$\lim_{J_B \to \infty} p_{\mathrm{d}}(\tilde{\mathbf{x}}_A) = q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)$. Hence

$$\lim_{J_B \to \infty} \chi^2\big(p_{\mathrm{d}}(\tilde{\mathbf{x}}_A), q_{\mathrm{d}}(\tilde{\mathbf{x}}_A)\big) = 0 \tag{28}$$

We conclude that Z_{d} can be estimated efficiently using the importance sampling estimator when, for $k \in \mathcal{B}_B$, the coupling parameters of the model are strong .

Similarly, one can show

$$\frac{L}{Z_{\mathrm{d}}^2}\mathrm{Var}[\hat{Z}_{\mathrm{Unif}}] = \chi^2\big(p_{\mathrm{d}}(\tilde{\mathbf{x}}_A), u_{\mathrm{d}}(\tilde{\mathbf{x}}_A)\big) \tag{29}$$

Let us assume that all the coupling parameters are constant, denoted by J. From (21), $\lim_{J \to \infty} p_{\mathrm{d}}(\tilde{\mathbf{x}}_A) = u_{\mathrm{d}}(\tilde{\mathbf{x}}_A)$, therefore, uniform sampling can efficiently estimate Z_{d} in models at very low temperature.

B. Further remarks

We finish this section with a few additional remarks.

i) Given the partitioning, the computational complexity of our algorithms is $O(N)$ per sample, which makes the total computational complexity $O(NL)$.

ii) Successive applications of our algorithm can yield all the marginals of (2), as ratios of two partition functions, with complexity $O(N^2L)$.

iii) The choice of partitioning in the dual NFG is arbitrary, as long as $\tilde{\mathbf{X}}_B$ forms a spanning tree in Fig. 3. From (28), a good heuristic strategy for partitioning in models with a mixture of strong and weak couplings is to include edges associated with stronger couplings in $\tilde{\mathbf{X}}_B$.

iv) If the couplings are relatively strong, we can employ annealed importance sampling [28] (see Appendix II). In the mid-temperature regime, the performance of MC methods in the dual NFG should be compared to the performance of MC methods applied directly to the primal NFG, as in Fig. 1.

v) For infinite-size models (i.e., as $N \to \infty$), we need to analyze the variance of estimating $\ln(Z)/N$, because Z itself becomes unbounded. For more details, see [13].

vi) Incidentally, the dual NFG gives an alternative derivation of the known analytical solution to the partition function of the one-dimensional (1D) Potts model (see Appendix III)

V. NUMERICAL EXPERIMENTS

In this section, we first consider tractable models amenable for exact computation and then apply our MC methods to larger instances, for which the exact computation is no longer feasible.

Tractable models: We consider instances of 8×8 3-state Potts models with periodic boundaries, the largest size solvable via the junction tree algorithm implemented in [29]. For comparison, we compute the relative error in estimating the log partition function, i.e., $|\ln Z - \ln \hat{Z}|/\ln Z$, vs. the coupling parameter, for the importance sampling and uniform sampling algorithms in the dual NFG, and the belief propagation (BP), the generalized BP (GBP), and the tree expectation propagation (TREEEP) algorithms in the primal NFG. These algorithms are the state of the art deterministic methods that perform best in our set-up among all the variants implemented in [29].

In GBP (GBPLoop), one cluster is used for each cycle of length four, and one for each maximal factor. In importance sampling and uniform sampling algorithms, we average over 50 trials, each with $L = 10^8$ samples, which takes about two minutes on a 2GHz Intel Xeon CPU.

Fig. 5 (left) shows the results for grids with constant coupling. At relatively high temperature (for $J < 1$), MC methods in the dual NFG perform poorly, while GBP gives the best results. In contrast, for $J > 1$, importance sampling and uniform sampling both perform well. In particular, importance sampling outperforms by more than an order of magnitude the second best approaches – uniform sampling and TREEEP (which perform comparably).

The effect of having mixed couplings is illustrated in Fig. 5 (middle), where for $k \in \mathcal{B}_B$ we set $J_k = 2$, and for $k \in \mathcal{B}_A$ couplings are fixed according to the value on the x-axis. We observe that importance sampling outperforms all the other approaches in a wider range of couplings (for $J > 0.3$), while BP and GBP give very poor results.

In Fig. 5 (right), we show the effect of having a constant external field $H = 0.1$ (see Appendix I) and mixed couplings as in Fig. 5 (middle). In this case, importance sampling performs better than the deterministic algorithms at very low temperature (for $J > 1.5$) while GBP and uniform sampling perform poorly.

Large grids: We show simulation results for the log partition function per site, i.e., $\frac{1}{N} \ln Z$, vs. the number of samples for one representative instance of the model with bond-dependent couplings. In other words, couplings are drawn
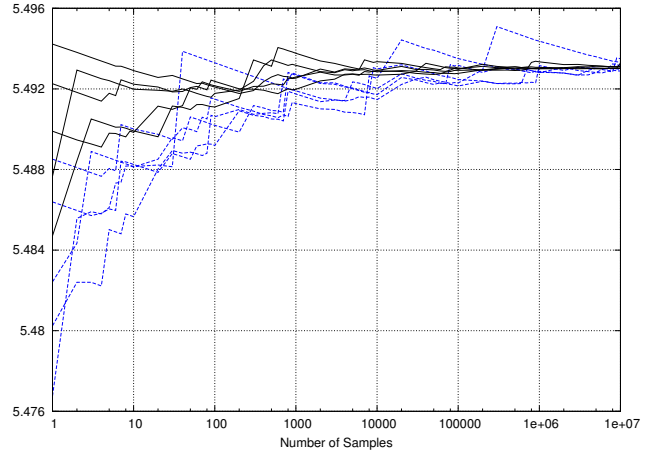

Fig. 6: Estimated $\ln(Z)/N$ vs. number of samples using IMP (solid black lines) and UNI (dashed blue lines) in the dual NFG of a 40×40 3-state Potts model with $J_k \sim \mathcal{U}[2.5, 3.0]$ for $k \in \mathcal{B}_A$ and $J_k \sim \mathcal{U}[2.5, 3.0]$ for $k \in \mathcal{B}_B$.
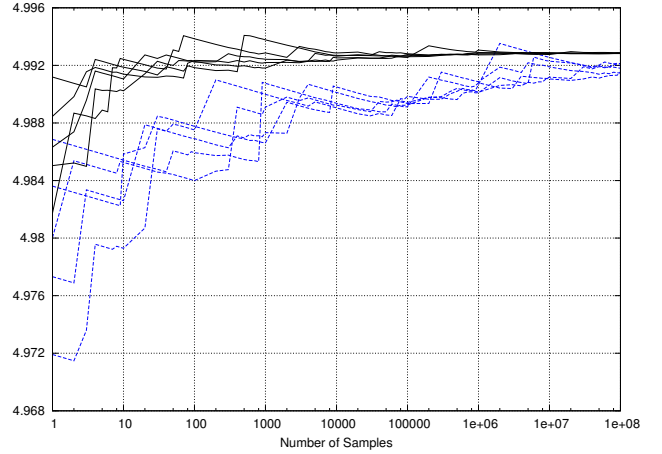

Fig. 7: Everything as in Fig. 6; but with $J_k \sim \mathcal{U}[2.0, 2.5]$ for $k \in \mathcal{B}_A$.

randomly; but remain fixed in the simulations[2].

We first consider a 3-state Potts model of size $N = 40 \times 40$, and fix $J_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}[2.5, 3.0]$ for $k \in \mathcal{B}_B$. For $k \in \mathcal{B}_A$, we set $J_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}[2.5, 3.0]$ and $J_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}[2.0, 2.5]$ in the first and in the second experiments, resp. Simulation results obtained from IMP (solid black lines) and UNI (dashed blue lines) in the dual NFG are shown in Figs. 6 and 7. For very large coupling parameters, convergence of the algorithms is comparable, see Fig. 6. However, for a wider range of couplings, uniform sampling has issues with slow convergence, see Fig. 7.

We then consider a 4-state Potts model with size $N = 30 \times 30$, and fix $J_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0.75, 2.25]$ for $k \in \mathcal{B}_A$. We set $J_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}[2.25, 3.25]$ in the third experiment (see Fig. 8) and $J_k \overset{\text{i.i.d.}}{\sim} \mathcal{U}[3.25, 4.25]$ in the fourth experiment (see Fig. 9), for

---

[2]In the jargon of statistical physics, estimating quantities for a fixed set of couplings (generated according to some distribution) is called the "quenched average". For more details, see [30, Chapter 2].
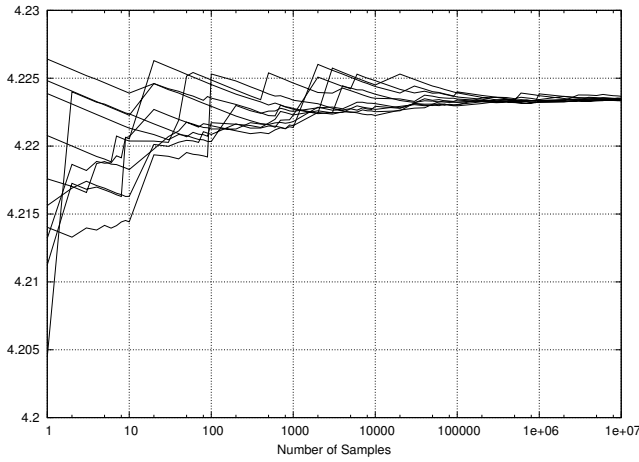
Fig. 8: Estimated $\ln(Z)/N$ vs. number of samples using IMP in the dual NFG of a 30×30 4-state Potts model with $J_k \sim \mathcal{U}[0.75, 2.25]$ for $k \in \mathcal{B}_A$ and $J_k \sim \mathcal{U}[2.25, 3.25]$ for $k \in \mathcal{B}_B$.
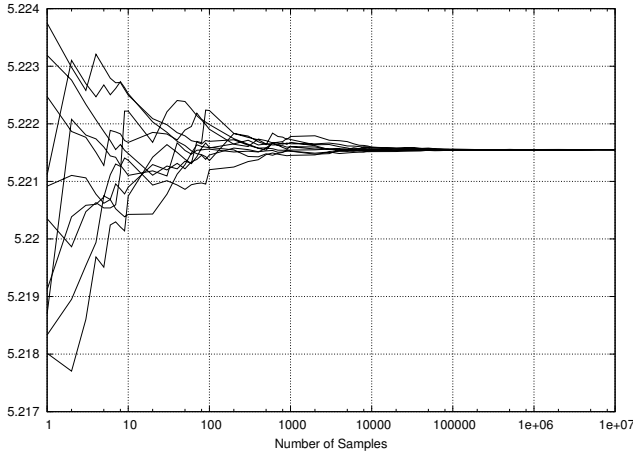

Fig. 9: Everything as in Fig. 8; but with $J_k \sim \mathcal{U}[3.25, 4.25]$ for $k \in \mathcal{B}_B$.

$k \in \mathcal{B}_B$. Simulation results obtained from IMP in the dual NFG are shown in Figs. 8 and 9. The estimated $\ln(Z)/N$ are about 4.223 and 5.2215, resp. In accordance with our analysis in Section IV-A, convergence of the importance sampling algorithm improves as J_k becomes stronger for $k \in \mathcal{B}_B$. Remarkably, for this size of grid and with such strong couplings, as in Fig. 9, good convergence is achieved using only a few thousands samples.

VI. CONCLUSION

We presented an importance sampling algorithm in the dual NFG of the 2D Potts model to estimate the partition function. In contrast to MC methods in the primal domain, convergence of our algorithm improves when the coupling parameters (on only a subset of the edges) become stronger. Furthermore, the method outperforms the state of the art deterministic algorithms in the primal domain, as well as uniform sampling methods in the dual domain, by more than an order of magnitude in terms of accuracy.

Since the dual NFG retains the topology of the primal NFG, we expect our algorithm to perform well in graphical models with arbitrary topology – especially at low temperature. Extending our methods to more general settings and applying variational inference in the dual NFG are the focus of our current research.

APPENDIX I
THE 2D POTTS MODEL IN AN EXTERNAL FIELD

In this section, we describe an importance sampling algorithm to estimate the partition function of the 2D q-state Potts model in the presence of an external field.

In this case, the energy of a configuration $\mathbf{x}$ is given by the Hamiltonian

$$\mathcal{H}(\mathbf{x}) = - \sum_{(k,\ell) \in \mathcal{B}} J_{k,\ell} \cdot [x_k = x_\ell] - \sum_{m=1}^{N} H_m \cdot [x_m = 0] \quad (30)$$

where the real coupling parameter $J_{k,\ell}$ controls the strength of the interaction between (x_k, x_ℓ) and the real parameter H_m corresponds to the presence of an external field.

In (30), we have assumed that the external field H_m, applies only if $x_m = 0$. The Hamiltonian can be defined in other ways, e.g., the field can apply when $x_m = 1$ or when x_m is in more than one state.

As in (3), for each adjacent pair (x_k, x_ℓ), we let

$$\kappa_{k,\ell}(x_k, x_\ell) = e^{J_{k,\ell} \cdot [x_k = x_\ell]} \quad (31)$$

and for each x_m, let $\tau_m : \mathcal{X} \to \mathbb{R}_{\geq 0}$

$$\tau_m(x_m) = e^{H_m \cdot [x_m = 0]} \quad (32)$$

We then define $f : \mathcal{X}^N \to \mathbb{R}_{\geq 0}$ as

$$f(\mathbf{x}) \triangleq \prod_{(k,\ell) \in \mathcal{B}} \kappa_{k,\ell}(x_k, x_\ell) \prod_{m=1}^{N} \tau_m(x_m) \quad (33)$$

The NFG of the 2D Potts model in an external field (i.e., the factorization in (33)) is shown in Fig. 10, where the unlabeled normal-size boxes represent factors (31) and the small boxes represent factors (32).

Again, we are interested in estimating the partition function Z, which is expressed as

$$Z = \sum_{\mathbf{x} \in \mathcal{X}^N} f(\mathbf{x}) \quad (34)$$

The dual NFG of the 2D Potts model is illustrated in Fig. 11, where factors attached to each equality indicator factor are as in (11), and factors attached to each mod_q indicator factor are given by

$$\lambda_m(\tilde{x}_m) = \begin{cases} (e^{H_m} + q - 1)/q, & \text{if } \tilde{x}_m = 0 \\ (e^{H_m} - 1)/q, & \text{otherwise,} \end{cases} \quad (35)$$
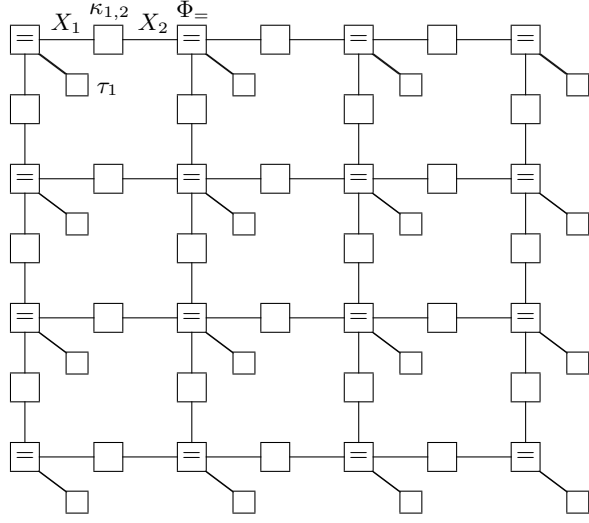
Fig. 10: NFG of the 2D Potts model in an external field, where unlabeled normal-size boxes represent factors (31), small boxes represent factors (32), and boxes containing "=" symbols are equality indicator factors.
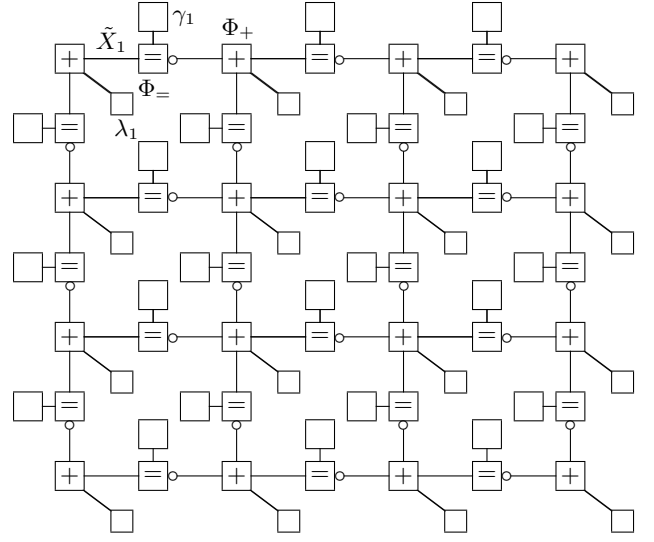

Fig. 11: Dual NFG of the 2D Potts model, where the unlabeled normal-size boxes represent (11), unlabeled small boxes represent (35), and the small circles denoted by "∘" are sign inverters.

which is the 1D DFT of (32).

We focus on ferromagnetic models (i.e., $J_k \geq 0$) in a nonnegative external field (i.e., $H_m \geq 0$), therefore, (35) will be nonnegative.

We can design an importance sampling algorithm in Fig. 11 to compute an estimate of Z_d. Again, we need to partition $\tilde{\mathbf{X}}$ into $\tilde{\mathbf{X}}_A$ and $\tilde{\mathbf{X}}_B$, with the condition that $\tilde{\mathbf{X}}_B$ is a linear combination (involving the XOR factors) of $\tilde{\mathbf{X}}_A$. Indeed, the same partitioning presented in Fig. 4 is applicable. We only need to add edges (variables) representing the external field to the thin edges, i.e., to $\tilde{\mathbf{X}}_A$.

For a valid configuration $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_A, \tilde{\mathbf{x}}_B)$, we suppose $\tilde{\mathbf{x}}_A = (\tilde{\mathbf{y}}, \tilde{\mathbf{z}})$, where $\tilde{\mathbf{y}}$ contains all the thin edges attached to the small unlabeled boxes (which represent variables involved in factors (35)), and $\tilde{\mathbf{z}}$ contains all the variables associated with the thin bonds (which represent variables involved in factors (11)).

We will need the following lemma.

Lemma 1. If $\tilde{\mathbf{x}}$ is a valid configuration in the dual NFG, then

$$\sum_{m=1}^{N} \tilde{y}_m = 0 \qquad (36)$$

Proof. We consider the sum of all the N components of $\tilde{\mathbf{y}}$ as, $c = \sum_{m=1}^{N} \tilde{y}_m$. Each mod_q indicator factor imposes the constraint that all its incident variables sum to 0. In c, each $\tilde{y}_m$ can thus be expanded as the sum of the corresponding variables associated with the bonds. Furthermore, the variables on the bonds each appear twice in this expansion, once with a positive sign and once with a negative sign. We conclude that $c = 0$. ∎

Lemma 1 asserts that in a valid configuration $\tilde{\mathbf{x}}$ there is a linear dependency among the components of $\mathbf{y}$. Therefore,

we need to exclude one of the edges attached to the small unlabeled boxes from $\tilde{\mathbf{X}}_A$. Without loss of generality, we assume that the excluded edge is involved in $\lambda_N(\cdot)$.

We then define

$$\Psi(\tilde{\mathbf{x}}_A) \triangleq \prod_{k \in \mathcal{B}_A} \gamma_k(\tilde{x}_k) \prod_{m=1}^{N-1} \lambda_m(\tilde{x}_m) \qquad (37)$$

$$\Lambda(\tilde{\mathbf{x}}_B) \triangleq \lambda_N(\tilde{x}_N) \prod_{k \in \mathcal{B}_B} \gamma_k(\tilde{x}_k) \qquad (38)$$

We use the following auxiliary probability mass function in our importance sampling algorithm

$$q_d(\tilde{\mathbf{x}}_A) = \frac{\Psi(\tilde{\mathbf{x}}_A)}{Z_{q_d}} \qquad \forall \tilde{\mathbf{x}}_A \in \mathcal{X}^{|\mathcal{B}_A|} \qquad (39)$$

where the partition function Z_{q_d} is analytically available as

$$Z_{q_d} = \sum_{\tilde{\mathbf{x}}_A} \Psi(\tilde{\mathbf{x}}_A) \qquad (40)$$

$$= q^{|\mathcal{B}_A| + N - 1} \cdot \exp\left(\sum_{k \in \mathcal{B}_A} J_k + \sum_{m=1}^{N-1} H_m \right) \qquad (41)$$

The product form of (37) indicates that two separate subroutines are required at each iteration ℓ, in order to draw $\tilde{\mathbf{x}}_A^{(\ell)} = (\tilde{\mathbf{y}}^{(\ell)}, \tilde{\mathbf{z}}^{(\ell)})$ according to (39); one to draw the $\tilde{\mathbf{y}}^{(\ell)}$-part and the other to draw the $\tilde{\mathbf{z}}^{(\ell)}$-part.

Drawing the $\tilde{\mathbf{z}}^{(\ell)}$-part can be done using Algorithm 1. To draw the $\tilde{\mathbf{y}}^{(\ell)}$-part, we apply the following.

In line 3, $\left(1 + (q-1)e^{-H_m}\right)/q$ is $\lambda_m(0)/\sum_{t=1}^{q-1} \lambda_m(t)$, and setting $\tilde{y}_N^{(\ell)}$ in line 9 is done according to (36).

After generating $\tilde{\mathbf{x}}_A^{(\ell)} = (\tilde{\mathbf{y}}^{(\ell)}, \tilde{\mathbf{z}}^{(\ell)})$, we update $\tilde{\mathbf{x}}_B^{(\ell)}$ in order to create a valid configuration $\tilde{\mathbf{x}}^{(\ell)} = (\tilde{\mathbf{x}}_A^{(\ell)}, \tilde{\mathbf{x}}_B^{(\ell)})$. The samples are then used to compute an estimate of Z_d. MC estimates of

Algorithm 3 Drawing the $\tilde{\mathbf{y}}^{(\ell)}$-part

1: *draw* $u_1^{(\ell)}, u_2^{(\ell)}, \ldots, u_N^{(\ell)} \overset{\text{i.i.d.}}{\sim} \mathcal{U}[0,1]$
2: **for** $m = 1$ **to** $N - 1$ **do**
3:     **if** $u_m^{(\ell)} < \dfrac{1 + (q-1)e^{-H_m}}{q}$ **then**
4:       $\tilde{y}_m^{(\ell)} = 0$
5:     **else**
6:       *draw* $\tilde{y}_m^{(\ell)}$ *randomly in* $\{1, 2, \ldots, q-1\}$
7:     **end if**
8: **end for**
9: *set* $\tilde{y}_N^{(\ell)} = \sum_{m=1}^{N-1} \tilde{y}_m^{(\ell)} \bmod q$

Z_{d} are then used to estimate Z via the normal factor graph duality theorem.

APPENDIX II
ANNEALED IMPORTANCE SAMPLING IN THE DUAL NFG

Annealed importance sampling is used in the primal NFG by moving from a tractable (high temperature) distribution to the target distribution via a sequence of intermediate distributions.

We briefly discuss how to employ annealed importance sampling in the dual NFG to estimate the partition function, when J_k is relatively strong for $k \in \mathcal{B}_B$. For simplicity, let us assume that the coupling parameters associated with the thin edges and the coupling parameters associated with the thick edges are constant, denoted by J_A and J_B, respectively.

We thus denote the partition function by $Z_{\mathrm{d}}(J_A, J_B)$, and express $Z_{\mathrm{d}}(J_A, J_B)$ using a sequence of intermediate partition functions by varying J_B in V levels as

$$Z_{\mathrm{d}}(J_A, J_B) = Z_{\mathrm{d}}(J_A, J_B^{\alpha_V}) \prod_{v=0}^{V-1} \frac{Z_{\mathrm{d}}(J_A, J_B^{\alpha_v})}{Z_{\mathrm{d}}(J_A, J_B^{\alpha_{v+1}})} \qquad (42)$$

Here, unlike typical annealing strategies in the primal domain, $(\alpha_0, \alpha_1, \ldots, \alpha_V)$ is an increasing sequence with $1 = \alpha_0 < \alpha_1 < \cdots < \alpha_V$; see [28, Section 2].

If α_V is large enough, $Z_{\mathrm{d}}(J_A, J_B^{\alpha_V})$ can be estimated efficiently via our proposed importance sampling algorithm. As for the intermediate steps, a sampling technique that leaves the target distribution invariant (e.g., Metropolis-Hastings algorithm) is required at each level. These intermediate target probability distributions correspond to the intermediate partition functions. Also, V should be sufficiently large to ensure that intermediate target distributions are close enough.

APPENDIX III
PARTITION FUNCTION OF THE 1D POTTS MODEL

We use the normal factor graph duality theorem to compute the partition function of the 1D q-state Potts model of size N, and with periodic boundary conditions. The NFG of the model is shown in Fig. 12, where boxes represent factors as in 3 and edges represent the variables.

Following our approach in Section III and after some modifications, we can obtain the dual NFG of the 1D Potts model as in Fig. 13, where the unlabeled boxes represent factors as in (11), and boxes containing "=" symbols are equality indicator factors.
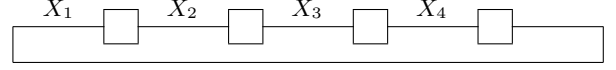

Fig. 12: The NFG of the 1D Potts model with $N = 4$ and with periodic boundary conditions.

Note that in the 1D case, the inverse Fourier transform of each equality indicator factor (with degree two) becomes a mod_q indicator factor with scale factor one.
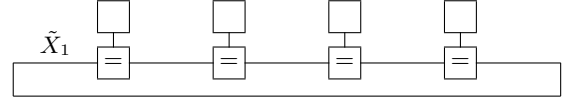

Fig. 13: Dual NFG of the 1D Potts model.

There are only q valid sequences in Fig. 13, namely, the all-zero sequence, the all-one sequence, etc. We can therefore compute Z_{d} directly as

$$Z_{\mathrm{d}} = \prod_{k=1}^{N}(e^{J_k} + q - 1) + (q-1)\prod_{k=1}^{N}(e^{J_k} - 1) \qquad (43)$$

Since all the local scale factors are one, $Z = Z_{\mathrm{d}}$.

Let us suppose that the coupling parameters of the model are constant, denoted by J. In thermodynamic limits of large N, we obtain

$$\lim_{N \to \infty} \frac{\ln Z}{N} = \ln(e^J + q - 1) \qquad (44)$$

In statistical physics, the transfer matrix method is usually employed to compute the partition function of classical 1D models. See [2] and [6, Chapter 5] for more details.

REFERENCES

[1] R. B. Potts, "Some generalized order-disorder transformations," *Proc. the Cambridge Philosophical Society*, vol. 48, pp. 106–109, 1952.

[2] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*. Dover Publications, 2007.

[3] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222–1239, 2001.

[4] F. Y. Wu, "Potts model and graph theory," *Journal of Statistical Physics*, vol. 52, pp. 99–112, 1988.

[5] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[6] J. M. Yeomans, *Statistical Mechanics of Phase Transitions*. Oxford University Press, 1992.

[7] R. Salakhutdinov and I. Murray, "On the quantitative analysis of deep belief networks," *Proc. of the 25th*

International Conference on Machine Learning, pp. 872–879, 2008.

[8] V. Gómez, H. J. Kappen, and M. Chertkov, "Approximate inference on planar graphs using loop calculus and belief propagation," *Journal of Mach. Learn. Res.*, vol. 11, pp. 1273–1296, 2010.

[9] L. A. Goldberg and M. Jerrum, "Approximating the partition function of the ferromagnetic Potts model," *Journal of the ACM*, vol. 59, pp. 1222–1239, 2012.

[10] R. M. Neal, "Probabilistic inference using Markov chain Monte Carlo methods," Department of Computer Science, University of Toronto, Tech. Rep. CRG-TR-93-1, 1993.

[11] K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics*. Springer, 2010.

[12] M. Molkaraie and H.-A. Loeliger, "Monte Carlo algorithms for the partition function and information rates of two-dimensional channels," *IEEE Trans. on Information Theory*, vol. 59, pp. 495–503, 2013.

[13] G. Potamianos and J. Goutsias, "Stochastic approximation algorithms for partition function estimation of Gibbs random fields," *IEEE Trans. on Information Theory*, vol. 43, pp. 1948–1965, 1997.

[14] I. Murray, D. Mackay, Z. Ghahramani, and J. Skilling, "Nested sampling for Potts models," *Advances in Neural Information Processing Systems*, pp. 947–954, 2005.

[15] R. H. Swendsen and J. S. Wang, "Nonuniversal critical dynamics in Monte Carlo simulations," *Physical review letters*, vol. 58, pp. 86–88, 1987.

[16] G. D. Forney, "Codes on graphs: normal realization," *IEEE Trans. on Information Theory*, vol. 47, pp. 520–548, 2001.

[17] ——, "Codes on graphs: duality and MacWilliams identities," *IEEE Trans. on Information Theory*, vol. 57, pp. 1382–1397, 2011.

[18] A. Al-Bashabsheh and Y. Mao, "Normal factor graphs and holographic transformations," *IEEE Trans. on Infor-*

mation Theory, vol. 57, pp. 752–763, 2011.

[19] M. Molkaraie and H.-A. Loeliger, "Partition function of the Ising model via factor graph duality," *Proc. IEEE International Symposium on Information Theory*, pp. 2304–2308, 2013.

[20] A. Al-Bashabsheh and Y. Mao, "On stochastic estimation of the partition function," *Proc. IEEE International Symposium on Information Theory*, pp. 1504–1508, 2014.

[21] D. E. Knuth, "Two notes on notation," *American Mathematical Monthly*, vol. 99, pp. 403–422, 1992.

[22] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Processing Mag.*, vol. 29, pp. 28–41, 2004.

[23] G. D. Forney and P. O. Vontobel, "Partition functions of normal factor graphs," *Information Theory and Applications Workshop*, 2011.

[24] R. N. Bracewell, *The Fourier Transform and Its Applications*. McGraw-Hill, 1999.

[25] M. Molkaraie, "An importance sampling scheme on dual factor graphs. I. models in a strong external field," `arXiv:1401.4912`, 2014.

[26] R. M. Karp and M. Luby, "Monte-Carlo algorithms for enumeration and reliability problems," *IEEE Annual Symposium on Foundations of Computer Science*, pp. 56–64, 1983.

[27] I. Csiszár and P. C. Shields, "Information theory and statistics: a tutorial," *Foundations and Trends in Communications and Information Theory*, vol. 1, pp. 417–528, 2004.

[28] R. M. Neal, "Annealed importance sampling," *Statistics and Computing*, vol. 11, pp. 125–139, 2001.

[29] J. M. Mooij, "libDAI: A free and open source C++ library for discrete approximate inference in graphical models," *Journal of Mach. Learn. Res.*, vol. 11, pp. 2169–2173, 2010.

[30] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing*. Oxford University Press, 2001.